Relatório Grupo de Trabalho

Deepfakes

Secretaria de Políticas Digitais
Secretaria de Comunicação Social
Presidência da República

PRESIDÊNCIA DA REPÚBLICA

Luiz Inácio Lula da Silva
Presidente da República

Geraldo Alckmin
Vice-Presidente

SECRETARIA DE COMUNICAÇÃO SOCIAL DA PRESIDÊNCIA DA REPÚBLICA

Sidônio Palmeira
Ministro de Estado da Secretaria de Comunicação Social da Presidência da República

Gabinete da Secretaria de Comunicação Social da Presidência da República

Samara Mariana de Castro
Chefe de Gabinete

Tiago Cesar dos Santos
Secretário-Executivo

SECRETARIA DE POLÍTICAS DIGITAIS

João Caldeira Brant Monteiro de Castro
Secretário de Políticas Digitais

Nina Fernandes dos Santos
Secretária Adjunta de Políticas Digitais

Lucas Abot Leffa
Assessor

Marina Giancoli Cardoso Pita
Diretora de Promoção da Liberdade de Expressão

André Parente Houang
Coordenador-Geral de Promoção da Diversidade e Pluralismo

Renata Negrelly Nogueira
Coordenadora-Geral de Liberdade de Expressão e Enfrentamento à Desinformação

Carolina Ofranti Sampaio
Coordenadora de Políticas para Liberdade de Expressão e Enfrentamento à Desinformação

Hasla de Paula Pacheco
Coordenadora de Promoção de Diversidade e Pluralismo

Brisa Queiroz Martins
Especialista em Financiamento e Execução de Programas e Projetos Educacionais

Pedro Ivo Badú de Oliveira
Diagramador

Sumário

Apresentação

O Grupo de Trabalho sobre *deepfakes* foi criado "com a finalidade de avaliar o impacto de conteúdos sintéticos relacionados a pessoas naturais, gerados ou manipulados por tecnologias de Inteligência Artificial, conhecidos como '*deepfakes*' e estudar o desenvolvimento de medidas para proteção de direitos fundamentais e contenção de fraudes e golpes". Ele foi instituído por meio da Portaria nº 1, de 22 de maio de 2025, pela Secretaria de Políticas Digitais da Secretaria de Comunicação Social da Presidência da República. O Grupo de Trabalho ficou encarregado de "identificar riscos e impactos da produção e disseminação de conteúdo sintético do tipo *deepfakes* aos direitos fundamentais e à proteção do patrimônio dos cidadãos e ao erário público", "realizar estudos técnicos acerca das possíveis medidas adotadas quanto ao fenômeno de produção e disseminação de *deepfakes* no Brasil e no exterior", e "recomendar diretrizes para a construção de políticas para contornar possíveis impactos danosos da produção e disseminação de ".

Integraram o grupo

João Brant;
Nina Santos;
Marina Pita;
André Houang;
Lucas Leffa;
Carolina Sampaio;
Lucia Maria Teixeira Ferreira;
Laura Schertel Mendes;
Diogo Cortiz;
Francisco Brito Cruz;
Tarcízio Silva;
Clara Keller;
Mariana Rielli;
Frederico Franco Alvim;
Anderson Schreiber;
Caitlin Mulholland;
Mariana Valente;
Filipe Medon;
Bia Barbosa;
João Vitor Archeg
Fernanda dos Santos R. Silva;
Matheus Soares;
Juliana Cunha
Thiago Tavares.

Apesar de o Grupo de Trabalho constituído não ter sido interministerial, alguns ministérios foram convidados a indicar representantes para participação, por conta de diálogos e trocas já estabelecidos anteriormente sobre a temática. O Ministério de Direitos Humanos e Cidadania, o Ministério das Mulheres, o Ministério da Justiça e Segurança Pública e a Advocacia-Geral da União responderam ao convite. Por conta da relevância da temática, também o Supremo Tribunal Federal (STF) foi convidado a indicar um especialista. Fruto desses convites, foram integrados ao GT:

Ísis Táboas (MM);
Raphael Souza (AGU);
Fabio Meireles (MDHC);
Julia Abad (MDHC);
Pedro de Barros Amaral (MJSP);
Victor Durigan (STF).

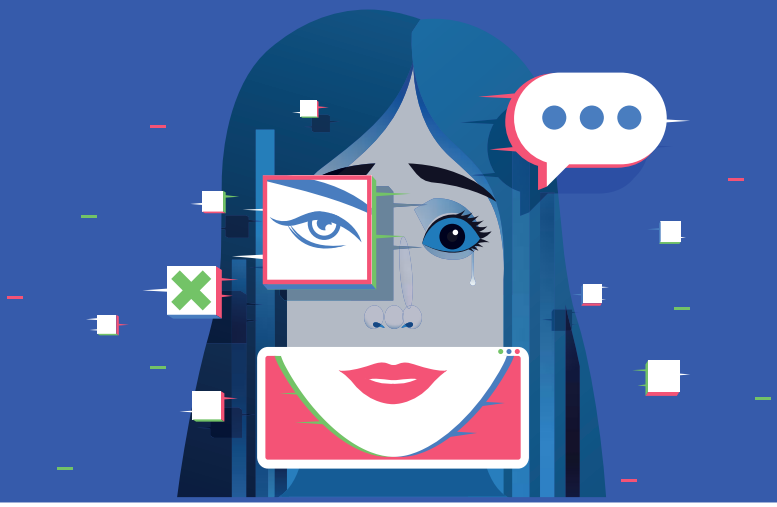
O grupo se reuniu quinzenalmente e, inicialmente, esteve focado na compreensão do conceito de *deepfakes*, em seus efeitos sobre diferentes grupos sociais e na sociedade. Também foram discutidos os possíveis mecanismos para a contenção dos efeitos deletérios das *deepfakes*, sem deixar de reconhecer o potencial positivo que sistemas de IA podem ter no futuro do Brasil.

Diante do cenário de tramitação, em Comissão Especial na Câmara dos Deputados, do Projeto de Lei (PL) n. 2338/2023, que visa estabelecer "normas gerais de caráter nacional para a governança responsável de sistemas de inteligência artificial (IA) no Brasil" (art. 1º), o Grupo de Trabalho passou então a focar seus esforços especificamente em desenvolver recomendações que pudessem ser incorporadas ao projeto - mas também analisou outras propostas regulatórias em discussão no legislativo.

A redação do **PL n. 2338/2023** aprovada pelo Senado Federal propunha regras sobre IA de alto risco e IA generativa, mas não continha dispositivos específicos para lidar com o fenômeno das *deepfakes*, tipo particular de conteúdo sintético que mimetiza a realidade a ponto de ser impossível de uma pessoa, desprovida de maiores recursos técnicos, distinguir sua natureza.

A partir do diagnóstico de que seria oportuno suprir essa lacuna no PL n. 2338/2023, o GT apresentou alternativas de alteração ou inclusão de dispositivos ao longo do PL. O relatório representa, assim, recomendações de alterações ao PL, independentes entre si. Alguns dos integrantes do GT participaram integralmente de todas as suas reuniões, enquanto outros participaram de forma intermitente. Este relatório é um esforço coletivo de reflexão e proposição acerca do tema de *deepfakes*. Este relatório foi elaborado a partir das discussões e propostas apresentadas no âmbito do GT, e não reflete necessariamente as opiniões de cada integrante do GT individualmente.

Considerando o esforço interministerial de diálogo sobre a matéria, especialmente os ministérios convidados a integrarem o Grupo de Trabalho, o relatório registra também as posições finais da Secretaria de Políticas Digitais em termos de recomendações para o legislativo.



1 Esferas sociais afetadas por *deepfakes*

Os modelos e aplicações de Inteligência Artificial (IA) generativa já se tornaram corriqueiros na vida dos brasileiros. Os usos dessas tecnologias são múltiplos, envolvendo aplicações na indústria, no gerenciamento de serviços e no cotidiano de milhões de brasileiros usuários de IA. Dentre as múltiplas aplicações desta tecnologia, a IA generativa tem ganhado destaque. Entendida como a IA destinada a gerar ou modificar texto, imagens, áudio, vídeo e código de software, a IA generativa já tem sido amplamente usada em diferentes indústrias, como nos setores musical, audiovisual e publicitário, e para usos corriqueiros em redes sociais.

Algumas das aplicações de IA generativa podem ser capazes de gerar as chamadas "*deepfakes*", conteúdos sintéticos que representam pessoas naturais, locais ou situações fictícias, porém similares à realidade. A possibilidade de produção de novos conteúdos a partir de mídias existentes ou de instruções específicas tem permitido tanto a geração de mídias entendidas como "*deepfakes*" com finalidades legítimas, protegidas pela liberdade de expressão, como sátiras e paródias, quanto também tem possibilitado a geração de imagens que podem atentar contra direitos difusos, coletivos, das pessoas retratadas e de terceiros.

As *deepfakes* afetam diferentes grupos e esferas sociais, sendo alguns grupos populacionais mais afetados do que os demais. Grupos hipervulneráveis e historicamente oprimidos estão mais sujeitos a serem impactados negativamente pelo uso abusivo de IA generativa. Frente ao uso crescente de ferramentas de IA generativa para a criação de *deepfakes*, reguladores e autoridades públicas de diversos países e jurisdições têm desenvolvido abordagens variadas para resguardar direitos frente a potenciais danos causados pelo uso dessas tecnologias de IA.

A seguir, passamos à análise de como alguns grupos e esferas sociais têm sido particularmente afetados negativamente pelas *deepfakes*. A análise apresentada não é exaustiva em termos de escopo nem de profundidade. Diversas outras esferas e grupos sociais têm sido afetados pelas *deepfakes* e os efeitos vão muito além do que os relatados neste documento. Há outras populações que podem ser fortemente afetadas por *deepfakes*, assim como as populações listadas a seguir podem ser afetadas de maneira distinta e mais severa do que a apresentada.

Proteção de crianças e adolescentes

Um primeiro ponto de sensibilidade diz respeito à geração de imagens de crianças a partir de sistemas de IA generativa. No Brasil, a Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), legislação referência para o tema, coloca ênfase na proteção de dados de crianças e adolescentes, mas os dispositivos da lei e sua aplicação podem ainda ser insuficientes, particularmente quanto às obrigações das pessoas jurídicas desenvolvedoras, distribuidoras e aplicadoras de sistemas de IA generativa quanto à geração de conteúdos de *deepfakes* envolvendo crianças e adolescentes.

Caso o conteúdo sintético gerado possibilite a sua associação à imagem, voz ou outro atributo de uma criança, é necessário que as proteções da própria LGPD, assim como do Estatuto da Criança e do Adolescente (Lei nº 8.069/1990) sejam aplicadas. No entanto, é necessária a criação de obrigações específicas para os desenvolvedores de sistemas de IA generativa, uma vez que a proteção da infância é prioridade absoluta, nos termos da Constituição brasileira (art. 227).

A questão é absolutamente relevante. Ferramentas de IA generativa capazes de gerar *deepfakes* de pessoas naturais têm sido crescentemente exploradas para a produção de material de abuso e exploração sexual de crianças e adolescentes. A *Internet Watch Foundation,* em pesquisa de 2023, identificou mais de 20.000 imagens geradas por IA publicadas em um fórum da web não indexada, em apenas um mês, sendo que mais de 3.000 retratavam atividades criminosas de abuso sexual infantil[1]. A organização aponta que as denúncias de imagens de exploração sexual de crianças e adolescentes geradas por IA mais que dobraram no último ano, aumentando de 199 em 2024 para 426 em 2025, com um aumento alarmante nas representações de bebês, sendo que imagens de crianças de 0 a 2 anos saltaram de 5 em 2024 para 92 em 2025.

O crescimento desses ilícitos tem sido exponencial: o National Center for Missing and Exploited Children (NCMEC) reporta que recebeu 67 mil denúncias de imagens de exploração sexual de crianças e adolescentes geradas por IA nos seis primeiros meses de 2024 e 485.000 no primeiro semestre de 2025, representando um aumento de 624%[2]. Entre 1º de janeiro e 31 de julho de 2025, 64% das denúncias anônimas registradas no Canal Nacional de Denúncias de Crimes Cibernéticos da ONG SaferNet foram sobre conteúdos digitais de abuso e exploração sexual infantil.

1 Internet Watch Foundation. How AI is being abused to create child sexual abuse imagery. IWF, Reino Unido, 2024. Disponível em: https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/. Acesso em: 19 jan. 2026

2 DAVIS, Patricia. Spike in online crimes against children a "wake-up call". Disponível em: https://www.missingkids.org/blog/2025/spike-in-online-crimes-against-children-a-wake-up-call. Acesso em: 19 de janeiro de 2026.

Particularmente preocupante é o surgimento de aplicativos de "nudificação" e ferramentas de deepfake que permitem a perpetradores manipular fotografias corriqueiras de redes sociais para criar conteúdo sexual explícito, frequentemente direcionado a crianças e adolescentes conhecidos pelos ofensores ou crianças famosas. O conteúdo gerado por IA tornou-se tão realista que é visualmente indistinguível de conteúdo de abuso e exploração sexual que não foi gerado por IA, mesmo para analistas treinados.

Como destacado anteriormente, esse fenômeno está crescendo em escala e sofisticação, superando as tentativas regulatórias atuais. No final de **2025 e início de 2026, após a incorporação de um módulo de geração de imagens, o chatbot de IA Grok, da xAI, disponível na plataforma de mídia social X (ex-Twitter), foi usado para "despir" mulheres e meninas, sem o seu consentimento ("undress nudes").** Usuários da plataforma X fizeram comentários em fotos públicas de pessoas reais e mencionavam o chatbot Grok, solicitando que ele criasse uma imagem da pessoa usando biquíni, muitas vezes em situações ou poses com conotação sexual. Esse comportamento rapidamente se transformou em uma tendência viral e, em diversas publicações na plataforma X, especialmente quando uma mulher aparecia, era comum encontrar usuários pedindo ao Grok que a "despissem" digitalmente.

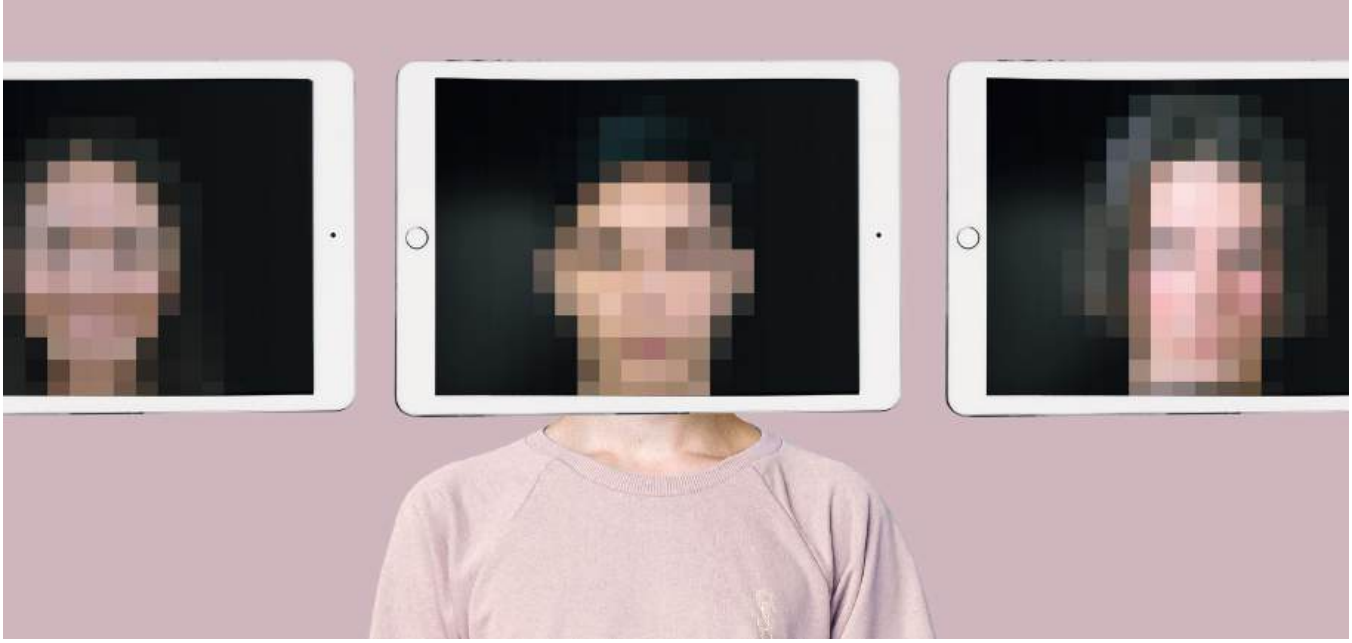
Essa produção automatizada e indiscriminada de conteúdo sexualizado não consentido pelo Grok, com especial impacto sobre crianças, adolescentes e mulheres, foi objeto de inúmeras denúncias ao redor do mundo, inclusive no Brasil[3], revelando, de forma cabal, a insuficiência de salvaguardas éticas e jurídicas no desenvolvimento e disponibilização pública de tecnologias baseadas em inteligência artificial.

A controvérsia envolvendo o Grok não representa um caso isolado, mas sim a manifestação de um fenômeno mais amplo e estrutural[4]. O episódio expôs a crescente proliferação de ferramentas baseadas em inteligência artificial voltadas à manipulação de imagens, bem como a expansão dos meios de acesso, compartilhamento e disseminação desses conteúdos. Em ambientes digitais como fóruns e comunidades online, como Reddit e Telegram, multiplicam-se discussões voltadas não apenas à exploração do potencial técnico dessas ferramentas, mas também à subversão deliberada das salvaguardas implementadas pelos próprios desenvolvedores, conhecidas como *guardrails.* Nessas redes, usuários compartilham estratégias com métodos de *jailbreak* dos modelos de IA, bem como maneiras para contornar restrições e mecanismos de segurança a fim de produzir conteúdos de cunho pornográfico, muitas vezes envolvendo imagens não consensuais e alvos vulneráveis. Esse cenário evidencia um déficit regulatório, além da necessidade urgente de responsabilização dos agentes envolvidos no desenvolvimento, hospedagem e difusão dessas tecnologias e dessas práticas ilícitas.

3-IDEC pede à ANPD suspensão do Grok no Brasil após imagens pornográficas geradas pela IA de Elon Musk. O Globo, Rio de Janeiro, 15 jan. 2026. Disponível em: https://oglobo.globo.com/economia/tecnologia/noticia/2026/01/15/idec-pede-a-anpd-suspensao-do-grok-no-brasil-apos-imagens-pornograficas-geradas-pela-ia-de-elon-musk.ghtml. Acesso em: 19 jan. 2026.

ANPD, MPF e Senacon recomendam que X impeça geração e circulação de conteúdos sexualizados indevidos por meio do Grok, 20 jan. 2026. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-mpf-e-senacon-recomendam-que-x-impeca-geracao-e-circulacao-de-conteudos-sexualizados-indevidos-por-meio-do-grok. Acesso em: 20 jan. 2026.

4-Use of AI to harm women has only just begun, experts warn. The Guardian, Londres, 14 jan. 2026. Disponível em: https://www.theguardian.com/technology/2026/jan/14/use-of-ai-to-harm-women-has-only-just-begun-experts-warn. Acesso em: 19 jan. 2026.



Pesquisas realizadas por organizações como o Institute for Strategic Dialogue[5] e o American Sunlight Project[6] identificaram a proliferação de aplicativos que produzem "nudificação" de imagens, muitos deles hospedados em lojas de aplicativos convencionais como as do Google e da Apple.

O uso da inteligência artificial para a produção e disseminação de imagens sexualizadas não consentidas, incluindo conteúdos envolvendo mulheres e crianças, representa uma grave violação aos direitos fundamentais à dignidade, privacidade, integridade psíquica e proteção contra a violência. No caso das crianças e adolescentes, a instrumentalização dessas tecnologias atinge diretamente a garantia de prioridade absoluta, prevista no artigo 227 da Constituição Federal, ao expô-los a formas sofisticadas de exploração e abuso, que escapam dos mecanismos tradicionais de controle e responsabilização. Além disso, esse fenômeno agrava as já persistentes violências de gênero ao digitalizar e automatizar práticas de objetificação, humilhação e silenciamento de meninas, ampliando o alcance do machismo estrutural por meio de ferramentas tecnológicas acessíveis e impunes. A lógica algorítmica, se não for regulada, pode transformar-se em vetor de discriminação e opressão, aprofundando desigualdades históricas e comprometendo os compromissos democráticos com os direitos humanos, especialmente os das populações mais vulnerabilizadas.

5 - PUGLIELLI, Chiara; CRAANEN, Anne. The ecosystem of nonconsensual intimate deepfake tools online. Institute for Strategic Dialogue (ISD), 2025. Disponível em: https://www.isdglobal.org/digital_dispatches/the-ecosystem-of-nonconsensual-intimate-deepfake-tools-online/. Acesso em: 19 jan. 2026

6-THE AMERICAN SUNLIGHT PROJECT. Meta's unsuccessful crackdown: how nonconsensual intimate imagery persists online. American Sunlight Project, 2025. Disponível em: https://americansunlight.substack.com/p/metas-unsuccessful-crackdown-how. Acesso em: 19 jan. 2026.

Mulheres e violência de gênero

As mulheres também são particularmente suscetíveis a terem seus direitos violados a partir da geração de *deepfakes*. Imagens, sons e vídeos gerados por sistemas de IA generativa podem reforçar vieses discriminatórios vigentes na sociedade, representando as mulheres em situações e posições que reproduzem e intensificam desigualdades históricas, explorando estereótipos misóginos e a objetificação do corpo feminino.

Especialmente preocupante para a proteção de direitos é a geração de *deepfakes* de mulheres em situação de nudez ou de ato sexual. Esses conteúdos são frequentemente gerados e difundidos sem consentimento das mulheres retratadas, com objetivo de difamá-las, violentá-las e humilhá-las, fato agravado por não haver sinalização de se tratar de um conteúdo gerado por IA. O uso da IA generativa para essa finalidade também ocorre em contextos políticos e eleitorais. Nos últimos anos, casos de *deepfakes* de candidatas em situação de nudez ou de ato sexual ganharam notoriedade no Brasil.[7]

A pesquisa "The State of *Deepfakes* 2023"[8], conduzida pela SecurityHero, aponta que, entre 2022 e 2023, a quantidade de pornografia deepfake criada aumentou em 464%. Em 2022, havia cerca de 3.725 vídeos de pornografia deepfake na internet, enquanto em 2023 esse número subiu para 21.019. A pornografia deepfake representa 98% de todos os vídeos deepfake online e 99% dos indivíduos alvos de pornografia deepfake são mulheres. No Brasil, o cenário é ainda mais preocupante. Estudos demonstram que o anúncio de ferramentas de geração de *deepnudes* é prática comum em buscadores, fragilizando sobretudo mulheres de grupos minorizados.[9]

Da perspectiva penal, importante salientar que, desde abril de 2025, está em vigor no Brasil a Lei n. 15.123/2025, que aumenta a pena de violência psicológica contra a mulher quando praticada com o uso de IA. Nesses casos, a punição, de seis meses a dois anos de prisão, é agravada em 50%. Ou seja, o legislativo brasileiro já avançou no processo de responsabilização das pessoas que realizam tal prática.

Porém, apesar da gravidade do tema e da existência de uma abordagem penal específica, não existe atualmente na legislação brasileira uma obrigação para desenvolvedores de sistemas de IA generativa para a proteção de mulheres e meninas que pudesse ser aplicada de forma preventiva. Diante desse cenário de risco aos direitos de mulheres, é importante que a regulação da IA traga obrigação de desenvolvedores de sistemas de IA generativa condicionarem a geração de *deepfakes* de pessoas em ato sexual ou de nudez a certas condições: que haja identificação do usuário que solicitou a geração do conteúdo sintético e a existência de dispositivo para sinalização, pelo usuário, de que houve consentimento da mulher retratada.

7 - CRUZ, M.; SANTOS, N.; CARREIRO, R.; NÓBREGA, L.; AMORIM, Gabriel. IA no primeiro turno: o que vimos até aqui. Salvador e São Paulo: Aláfia Lab & Data Privacy Brasil, 2024.
TAVARES, C.; VALENTE, M.; VILELA, C. Inputs to the 2026 Thematic Report on Gender Equality, the Digital Space and the Age of Artificial Intelligence- WGDAW. São Paulo: InternetLab, 2025.

8 - SECURITY HERO. The State of *Deepfakes*, 2023. Dipsonível em: https://www.securityhero.io/state-of-deepfakes/ . Acesso em: 19 de janeiro de 2026.

9 - RODRIGUES, Fernanda. Retratos da Objetificação de Mulheres Negras. In: SOUZA.,G.; SILVA, T. (orgs). Enfrentando *Deepfakes*, 2ªed. São Paulo: Desvelar, 2026

Racismo e *deepfakes* no Brasil

Os sistemas de IA generativa são geralmente desenvolvidos a partir de dados naturais, coletados e processados sem curadoria fina, logo enviesados em diversos aspectos. Adicionalmente, a literatura científica e o jornalismo mostram que os danos a grupos politicamente minorizados, como a população negra, é fruto da ausência de controle e testes amplos de qualidade para diferentes grupos. A produção de *deepfakes* de pessoas não-brancas pode reforçar estereótipos negativos, que invisibilizam a realidade e diversidade ou reforçam representações discriminatórias.

A automação desses vieses e a falta de controle amplia seu alcance, pois conteúdos racistas podem ser produzidos em larga escala, com aparência de veracidade e grande capacidade de disseminação. Ao mesmo tempo, sistemas de IA podem ser e já são deliberadamente desenvolvidos e aplicados para gerar conteúdos discriminatórios e incitar discriminação e violência contra pessoas racializadas, o que atenta contra os direitos desses cidadãos[10]. O combate a esse fenômeno exige, portanto, dispositivos legais que permitam o enfrentamento a vieses raciais e discriminatórios de sistemas de IA.

Há exemplos documentados de casos de vieses discriminatórios contra pessoas não-brancas. Um estudo de 2023 testou sistematicamente prompts em diferentes modelos e encontrou que estereótipos raciais persistiam consistentemente, com profissões de prestígio associadas a brancos e trabalhos manuais/serviços a pessoas não-brancas[11]. Pesquisa acadêmica publicada na prestigiada revista The Lancet aponta que o sistema Midjourney não apenas reproduz imagens discriminatórias de pessoas negras no contexto de saúde, como esforços de corrigir os vieses se mostraram infrutíferos[12]. Quanto a indivíduos de grande visibilidade pública, como jogadores de futebol, o uso de *deepfakes* tem sido operacionalizado sobretudo contra personalidades que se posicionam contra o racismo, a exemplo do uso da tecnologia contra o jogador Vini Jr[13]. A tendência corrobora descobertas já identificadas sobre o funcionamento do racismo online no país[14].

10 MCINTYRE, Niamh. "New AI video tools are fuelling violent racism on TikTok", The Bureau of Investigative Journalism", 16 Out. 2025, disponível em https://www.thebureauinvestigates.com/stories/2025-10-16/new-ai-video-tools-are-fuelling-racism-on-tiktok . Aceso em : 19 de janeiro de 2026.

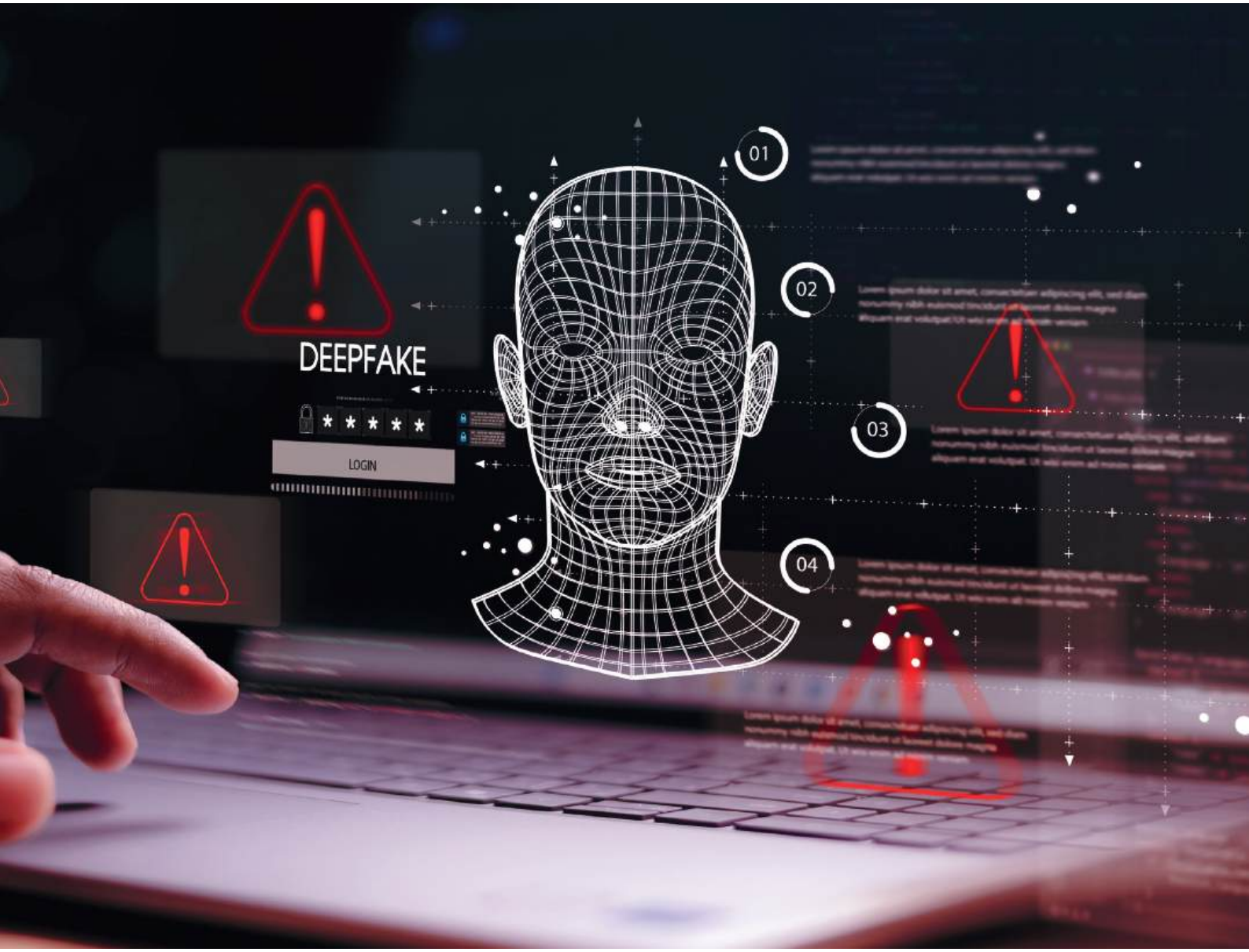
11 NICOLETTI, Leonardo; BASS, Dina. Humans are biased. Gejerative AI is even worse. Bloomberg Technology, 2023. Disponível em: https://www.bloomberg.com/graphics/2023-generative-ai-bias/ . Acesso em: 19 de janeiro de 2026.

12 GANZ , M. et al. The global health implications of deepfake technologies. The Lancet Global Health, v. 11, n. 10, 2023. Disponível em: https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(23)00329-7/fulltext. Acesso em: 19 jan. 2026.

13 - DANTAS, Glenda. Racismo Recreativo de Hoje. In: SOUZA.,G.; SILVA, T. (orgs). Enfrentando *Deepfakes*, 2ªed. São Paulo: Desvelar, 2026

14 TRINDADE, Luiz Valério. It is not that funny: Critical analysis of racial ideologies embedded in racialized humour discourses on social media in Brazil. 2018. Tese de Doutorado. University of Southampton.

Outra pesquisa que reuniu pesquisadores de renomadas universidades indica que sistemas de IA Generativa, ao receberem comandos ordinários, incluindo comandos que simplesmente mencionam características, ocupações ou objetos, descritores, produzem imagens estereotipadas. Por exemplo, identificam-se casos em que a solicitação de características básicas ou papéis sociais resulta em imagens que reforçam a branquitude como ideal, comandos referentes a ocupações resultam na amplificação de disparidades raciais e de gênero, e comandos relacionados a objetos resultam na reificação de normas norte-americanas. Os elementos discriminatórios estão presentes independentemente de os comandos mencionarem explicitamente identidade e linguagem demográfica ou evitarem tal linguagem. Além disso, os estereótipos persistem apesar das estratégias de mitigação; nem as tentativas dos usuários de contrapor estereótipos solicitando imagens com contra-estereótipos específicos, nem as tentativas institucionais de adicionar "salvaguardas" aos sistemas conseguiram prevenir a perpetuação de estereótipos[15].



15 - BIANCHI et al. Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23). Association for Computing Machinery, New York, NY, USA.

Confiança pública: fraudes, golpes e Democracia

As fraudes e golpes envolvendo as *deepfakes* têm se tornado cada vez mais frequentes, gerando também queda da confiança geral da população na tecnologia[16]. É comum que essas práticas se valham de táticas como a criação de *deepfakes* de figuras públicas anunciando produtos ou serviços; ou a criação de sites e anúncios falsos, inclusive usando marcas de órgãos ou campanhas do poder público ou de grandes empresas. Em junho de 2025, o Oversight Board de Meta publicou decisão em um caso em que uma deepfake produzida a partir da imagem do jogador de futebol Ronaldo estava sendo usada para direcionar pessoas a um jogo de azar online. Segundo a decisão do Overisght Board, "*Deepfakes* e promoções com *deepfakes* estão aumentando globalmente, incluindo aqueles envolvendo figuras públicas promovendo campanhas políticas fraudulentas e golpes financeiros. Relatórios destacam que muitos dos golpes financeiros no Brasil originados no Facebook, no Instagram e no WhatsApp envolvem conteúdo manipulado por IA[17]. " Novamente, obrigações de identificação do conteúdo gerado e de rastreabilidade podem ser essenciais para o enfrentamento ao uso de *deepfakes* para essas práticas criminosas e para a proteção de usuários e de consumidores.

Além de serem usadas em práticas criminosas e ilegais, as *deepfakes* desafiam as ideias de autenticidade, de fatos e de verdade compartilhada, afetando a base epistêmica para deliberação coletiva e, assim, as condições essenciais para a democracia. A sofisticação linguística e imagética e a capacidade de adaptação dos grandes modelos de linguagem representam uma ameaça sem precedentes à integridade informacional, pois permitem a produção automatizada e em larga escala de desinformação altamente persuasiva, com custos mínimos e sem necessidade de habilidades especializadas, potencialmente democratizando o acesso a ferramentas de manipulação em massa que anteriormente exigiriam recursos significativos de estados ou organizações bem financiadas. Adicionalmente, as principais ferramentas com infraestrutura global para fornecimento de serviços de deepfake são dominadas por um pequeno número de empresas.

Segundo o estudo "Generative AI and Democracy – Impacts and Interventions"[18], há quatro mecanismos pelos quais a IA generativa pode gerar efeitos sobre as eleições e a estabilidade democrática. Primeiramente, a IA pode ser usada para desinformação e desconfiança, influenciando no que as pessoas acreditam sobre políticos, problemas políticos e procedimentos eleitorais. A geração de *deepfakes* com políticos fazendo ou falando algo que não aconteceu, inclusive instruindo eleitores a não votarem, são exemplos de uso de IA generativa que se enquadram nesse primeiro mecanismo.

Além disso, a IA generativa pode ser empregada na geração de material abusivo, com o objetivo de afastar indivíduos ou grupos de se engajarem política ou eleitoralmente. Como exemplo desse mecanismo, no Brasil e em diversos países já são numerosos os casos de uso de IA generativa para produzir materiais que atacam diretamente candidatas mulheres e jornalistas, o que pode intimidá-las e afastá-las da vida pública e política[19].

Um terceiro mecanismo elencado pelo estudo britânico é o uso de ferramentas de IA generativa para a realização de cyber-ataques contra a infraestrutura eleitoral e campanhas

16 - BRAZIL FORUM. Relatório de pesquisa Brasil Forum UK. 2025. Disponível em: https://brazilforum.org/pesquisa/. Acesso em: 19 jan. 2026

17 - OVERSIGHT BOARD. Vídeo manipulado por IA promovendo jogos de azar. 5 de junho de 2025. Disponível em: https://www.oversightboard.com/decision/fb-o7ai7uax .

18 - SEGER, Elizabeth. Generative AI and Democracy – Impacts and Interventions. Demos, Abril de 2024.

19 - PEREIRA, Laura. A Banalização do Sintético. In: SOUZA.,G.; SILVA, T. (orgs). Enf rentandoDeepf akes, 2ªed. São Paulo: Desvelar, 2026

políticas. A IA generativa pode ser usada em golpes e roubos de identidade ou ainda na produção de softwares maliciosos. Na regulamentação eleitoral brasileira, em que pese a proibição estabelecida pelo § 1º do art. 9º-C da Resolução TSE nº 23.610/2019, conforme a redação dada pela Resolução nº 23.732/2024, que veda expressamente a criação e veiculação de *deepfakes* em campanhas eleitorais, a crescente disseminação de *deepfakes* representa um risco significativo, pois coloca em xeque a confiança do eleitorado na autenticidade das informações e na própria lisura do processo eleitoral[20].

Essas falsificações digitais, quando combinadas a outras tecnologias como bots de engajamento, chatbots, sistemas de microdirecionamento e algoritmos de recomendação, vêm sendo empregadas para manipular artificialmente a opinião pública, propagar desinformação em larga escala e distorcer o debate político, dificultando a distinção entre o que é real e o que é fabricado. Trata-se de um cenário que ameaça diretamente os pilares do processo democrático, exigindo respostas regulatórias eficazes e ações preventivas coordenadas entre instituições eleitorais, plataformas digitais e sociedade civil[21].

Nos últimos anos, e especialmente em 2024 - o chamado "super ano eleitoral" - e em 2025, a criação de *deepfakes* se tornou mais fácil, barata e sofisticada, resultando em sua proliferação em campanhas eleitorais em todo o mundo. Uma análise de cerca de 20 mil publicações relacionadas às eleições na Holanda, por exemplo, mostrou a presença crescente de conteúdos manipulados por IA, o que demonstra a fragilidade das defesas democráticas europeias contra a manipulação digital[22].

O uso político indiscriminado e sem regras dessas tecnologias representa uma ameaça concreta à integridade das democracias. *Deepfakes* realistas estão sendo usadas para desacreditar candidatos, difamar adversários e inflamar debates políticos nas redes sociais. Esse tipo de conteúdo, muitas vezes viralizado por algoritmos que priorizam engajamento, se espalha rapidamente antes que qualquer verificação seja possível — o que aumenta os riscos de confusão pública, polarização e perda de confiança nas instituições.

Nas eleições municipais brasileiras de 2024 foi possível ter uma amostra do poder da IA. Da perspectiva de danos, os casos de "deep nudes" — montagens digitais com nudez ou conteúdo sexual geradas por inteligência artificial — configuram uma nova forma de violência de gênero no ambiente eleitoral. Mulheres e grupos minorizados, já historicamente expostas a discriminação e ataques difamatórios, enfrentam agora um agravamento desse cenário, com a IA sendo usada como instrumento para potencializar práticas abusivas.

Apesar da cobertura jornalística sobre episódios envolvendo deep nudes, especialmente

20 - JUNQUILHO, Tainá Aguiar; SILVEIRA, Marilda de Paula; FERREIRA, Lucia Maria Teixeira; MENDES, Laura Schertel; OLIVEIRA, André Gualtieri de. (org.). Construindo consensos: deep fakes nas eleições de 2024 relatório das decisões dos TREs sobre deep fakes. Brasília: Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa: Laboratório de Governança e Regulação de Inteligência Artificial, 2024, E-book.

21 - FERREIRA, Lucia Maria Teixeira. A dimensão objetiva do direito fundamental à proteção de dados pessoais: perfilamento e microdirecionamento de propaganda político-eleitoral digital por provedores de aplicação de internet. Rio de Janeiro: Lumen Juris, 2025.

22 - HAECK, Pieter; HARTOG, Eva. The week that AI deepfakes hit Europe's elections. Politico. 31 out. 2025. Disponível em: https:// www.politico.eu/article/elections-europe-ai-deepfakes-social-media/ . Acesso em: 19 de janeiro de 2026.

com mulheres como principais vítimas, verificou-se na pesquisa do LIA/CEDIS/IDP [23] que é muito difícil combater essa prática no âmbito das ações judiciais eleitorais, o que pode refletir múltiplas situações, tais como: (i) lentidão nas investigações, que prejudica a atuação dos tribunais; (ii) dificuldades enfrentadas pelas vítimas para acionar a Justiça Eleitoral, enfraquecendo a visibilidade do problema; (iii) os efeitos colaterais da denúncia sobre a vida pessoal e a campanha política das vítimas; e (iv) o agravamento do ambiente de violência e intimidação.

Países como Estados Unidos, Irlanda, Holanda, Paquistão, Japão, Índia e Argentina já enfrentaram esse desafio nas eleições recentes, com a circulação de conteúdos falsos que dificultam a distinção entre fato e manipulação. Em democracias, isso resulta na erosão do consenso sobre os fatos básicos: escândalos reais podem ser descartados como *deepfakes*, enquanto *deepfakes* convincentes podem ser tomados como verdades. A autenticidade das evidências torna-se disputável, comprometendo o próprio funcionamento das instituições eleitorais.

Esse cenário é agravado pela vulnerabilidade do público. Em uma pesquisa do The Alan Touring Institute, embora 87% dos britânicos demonstrem preocupação com o impacto das *deepfakes* nas eleições, muitos não se sentem capazes de identificá-las, o que aumenta o risco de manipulação eleitoral direcionada e personalizada[24]. A IA torna esses conteúdos não apenas mais sofisticados, mas também mais persuasivos e emocionalmente impactantes.

O impacto não se limita às democracias liberais. Em regimes autoritários, a proliferação de *deepfakes* e teorias da conspiração pode legitimar a repressão e fortalecer o controle estatal, ao mesmo tempo em que alimenta o medo e a desinformação entre a população. Além disso, há o risco crescente de que a sociedade se fragmente ainda mais entre os que confiam nas tecnologias digitais e os que as rejeitam, acirrando divisões políticas, culturais e geracionais.

Diante desse cenário, é urgente que governos, sociedade civil, plataformas digitais e instituições internacionais avancem na construção de mecanismos regulatórios, normas éticas e estratégias educativas capazes de proteger a integridade dos processos democráticos sem sufocar a liberdade de expressão. Caso contrário, estaremos cada vez mais próximos de uma realidade na qual a verdade se torna irreconhecível, e a própria democracia, insustentável.

Por outro lado, os sistemas de IA Generativa podem ser amplamente usados para fins lícitos e apoiar a promoção do direito à liberdade de expressão, do direito à informação e à integridade da informação, além de apoiar os trabalhadores na realização de tarefas.

Em busca de um equilíbrio entre a necessidade de evitar os impactos negativos na integridade da informação sem ameaçar o exercício de direitos fundamentais no atual cenário tecnológico, diversos países e iniciativas de autorregulação têm resultado em obrigações e sistemas de sinalização e rastreabilidade do conteúdo gerado.

23 - JUNQUILHO, Tainá Aguiar; SILVEIRA, Marilda de Paula; FERREIRA, Lucia Maria Teixeira; MENDES, Laura Schertel; OLIVEIRA, André Gualtieri de. (org.). Construindo consensos: deep fakes nas eleições de 2024 relatório das decisões dos TREs sobre deep fakes. Brasília: Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa: Laboratório de Governança e Regulação de Inteligência Artificial, 2024, E-book.

24 - DENNEHY, Fiona. 9 in 10 concerned about deepfakes affecting election results. The Alan Turing Institute. 02 julho 2024. https://www.turing.ac.uk/news/9-10-concerned-about-deepfakes-affecting-election-resul. Acesso em: 19 de janeiro de 2026.

2 Recomendações para enfrentamento a *deepfakes*

Projeto de Lei n. 2.338/2023

Está em tramitação no Congresso Nacional o Projeto de Lei n. 2.338/2023, que "dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana". O projeto de lei (PL) propõe a criação de um arcabouço para promoção do fomento para a regulação de IA. O texto aprovado pelo Senado Federal, ao final de 2024, e enviado para apreciação na Câmara dos Deputados trata de sistemas de IA generativa na seção V do capítulo IV, intitulada "Das Medidas de Governança para Sistemas de Inteligência Artificial de Propósito Geral e Generativa".

Na seção de IA de propósito geral e IA generativa, a redação do PL 2338/2023 aprovado pelo Senado Federal cria obrigações ao desenvolvedor desses sistemas, tais como a realização de avaliação preliminar (art. 29), a garantia de certos requisitos no caso de sistema de IA generativa com risco sistêmico (art. 30), um dever de cooperação com os demais agentes de IA (art. 32) e a possibilidade de simplificação das obrigações em determinadas hipóteses (art. 33).

A despeito dessa seção específica para IA generativa, o texto aprovado pelo Senado Federal não traz medidas regulatórias específicas para prevenir a violação de direitos relacionada à criação de "*deepfakes*". No capítulo de sistemas de alto risco há apenas menção à necessidade de inserir dados para verificação de autenticidade e de proveniência de conteúdo sintético gerado por IA (art. 19). Porém, o próprio texto não trata sistemas de propósito geral com capacidade generativa como de alto risco, de forma que os principais geradores de "*deepfakes*" não estariam sujeitos a tal obrigação.

Assim, o Grupo de Trabalho "*Deepfakes*" considerou oportuno produzir propostas de redação com o foco em proteção de direitos por meio da criação de obrigações às pessoas jurídicas responsáveis por sistemas de IA Generativa no Projeto de Lei 2338/2023 e enviar as contribuições aos integrantes da Comissão Especial da Câmara dos Deputados, onde a matéria tramita. Diversos parlamentares da Câmara dos Deputados apontaram a necessidade de sanar essa lacuna, com a inserção de dispositivos específicos à disciplina de *deepfakes*[25].

A seguir, apresentamos uma breve discussão das problemáticas, indicação e posicionamento do Grupo de Trabalho e a proposta final adotada pela Secretaria de Políticas Digitais, a partir das discussões realizadas, em diálogo com os demais ministérios envolvidos na avaliação da matéria, bem como com a Casa Civil e a Secretaria de Relações Institucionais da Presidência da República.

Definição de Deepfake para fins da lei

A redação do PL 2.338/2023 aprovado pelo Senado Federal apresenta em seu art. 4º as definições adotadas pelo projeto de lei. Entre as definições da lei, não consta atualmente uma definição específica para "*deepfakes*". No entanto, o projeto define "conteúdos sintéticos" como "informações, tais como imagens, vídeos, áudio e texto, que foram significativamente modificadas ou geradas por sistemas de IA" (art. 4º, inciso XXI).

A definição de "conteúdo sintético" é ampla e já abarca *deepfakes*, de forma que as regras previstas para esse tipo também se aplicariam para *deepfakes*. Porém, ainda que a definição atual permita a aplicação dos dispositivos, seria possível fazer alterações para aprimorar a definição existente ou, caso se considere mais apropriado, também seria possível incluir uma definição própria específica para *deepfakes*.

O Grupo de Trabalho de *Deepfakes* da Secretaria de Políticas Digitais explorou os dois caminhos:

Sugestão A: Não incluir nova definição específica de *deepfakes*. Tratar dos problemas incluindo novos dispositivos a partir da definição de conteúdo sintético, com pequeno ajuste de redação.

> *XXI – conteúdos sintéticos: informações, tais como imagens, vídeos, áudio e texto que foram significativamente modificadas ou geradas por sistemas de IA;*

Sugestão B: inclusão de novo inciso com definição de "conteúdo sintético assemelhado à realidade", evitando o anglicismo de "*deepfakes*".

> *XXXI - conteúdo sintético assemelhado à realidade: conteúdo sintético que tenha aparência de pessoas, objetos, locais, entidades ou acontecimentos reais, e que parecem autênticos.*

A opção de não incluir novo inciso com definição de *Deepfakes* foi a escolhida pela Secretaria de Políticas Digitais para orientar sua posição nos diálogos interministeriais porque implicaria em menor interferência em texto já aprovado no Senado Federal. Adicionalmente, o conceito de conteúdo sintético poderia ser melhor circunscrito em cada artigo, a depender da necessidade.

25 Nas redações apresentadas a seguir, trechos taxados e em vermelho são propostas de exclusão do texto aprovado pelo Senado Federal, enquanto textos em verde são propostas de inserção.

Direito da pessoa ou grupo afetado por sistema de IA

A redação aprovada pelo Senado Federal traz, no artigo 5º, como direitos referentes a pessoa ou grupo afetado por sistema de IA, o direito à informação, à privacidade e à não discriminação.

Frente ao fenômeno das *deepfakes*, as pessoas ou grupos afetados por sistemas de IA também devem poder resguardar outros direitos frente a sistemas de IA generativa. Para tal, poderiam ser incluídos parágrafos ou incisos prevendo também os direitos de personalidade em face a sistemas de IA. Para evitar o abuso do direito de imagem, o Grupo de Trabalho considerou importante atentar para a necessidade de sopesamento desse direito com o direito à liberdade de expressão e à livre criação de sátiras e paródias.

A redação aprovada no Senado Federal do Projeto de Lei analisado, no art. 66, trata de direitos de personalidade:

> ***Art. 66.*** *A utilização de conteúdos de imagem, áudio, voz ou vídeo que retratem ou identifiquem pessoas naturais pelos sistemas de inteligência artificial deverá respeitar os direitos da personalidade, na forma prevista no Código Civil e na legislação pertinente.*

O artigo foi incluído em seção específica sobre direitos autorais, com objetivo legítimo de proteger direitos de artistas e intérpretes. A redação ampla, caso inserida em outro capítulo do projeto de lei, poderia proteger as pessoas ou grupos afetados por sistemas de IA, de forma ampla, quanto ao fenômeno das *deepfakes*. Mas, ainda não contemplaria a necessidade de ponderação ante o direito do legítimo direito à expressão livre.

Assim, ante a preocupação manifestada pelos integrantes do GT, propõe-se a inclusão de dispositivo no art. 5º, com adequação:

Sugestão A - inclusão de inciso IV ao art.5º. A inserção do termo "e seus produtos", de forma a sinalizar o direito de conhecimento acerca da qualidade sintética de uma produção de imagem, vídeo, áudio ou texto utilizando sistemas de IA. Além disso, foi proposto um novo inciso ao art. 5º do projeto de lei que equipara o direito à imagem aos demais direitos já previstos no Projeto de lei. O aprimoramento da redação do primeiro inciso e do §1º reforçaria o direito à informação e, portanto, a integridade da informação, o que é importante para lidar com os desafios associados às *deepfakes*.

> *Art. 5º. A pessoa ou grupo afetado por sistema de IA, independentemente do seu grau de risco, tem os seguintes direitos, a serem exercidos na forma e nas condições descritas neste Capítulo:*
>
> *I - direito à informação quanto às suas interações com sistemas de IA e seus produtos, de forma acessível, gratuita e de fácil compreensão, inclusive sobre caráter automatizado da interação, exceto nos casos em que se trate de sistemas de IA dedicados única e exclusivamente à cibersegurança e à ciberdefesa, conforme regulamento (NR);*
>
> *II – direito à privacidade e à proteção de dados pessoais, em especial os direitos dos titulares de dados nos termos da Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais), e da legislação pertinente;*

III – direito à não discriminação ilícita ou abusiva e à correção de vieses discriminatórios ilegais ou abusivos, sejam eles diretos ou indiretos.

IV – direito à imagem, nos termos da Lei nº 10.406, de 10 de janeiro de 2002 (Código Civil), inclusive quanto à geração ou modificação de imagem, áudio ou vídeo por IA generativa, resguardado o direito à liberdade de expressão; (NOVO INCISO)

§ 1º A informação referida no inciso I do caput deste artigo será fornecida com o uso de ícones ou símbolos uniformizados facilmente reconhecíveis e legíveis por máquina, sem prejuízo de outros formatos.

§ 2º Os sistemas de IA que se destinem a grupos vulneráveis deverão, em todas as etapas de seu ciclo de vida, ser transparentes e adotar linguagem simples, clara e apropriada à idade e à capacidade cognitiva, e ser implementados considerando o melhor interesse desses grupos.

Sugestão B – inserção de parágrafo ao art. 5º, a partir do atual art. 66: a inserção de um parágrafo ao texto do art. 5º não equipara o direito à imagem aos demais direitos previstos nos incisos do art. 5º, mas aproxima a proposta de texto da redação do art. 66 do relatório aprovado no Senado.

Art. 5º A pessoa ou grupo afetado por sistema de IA, independentemente do seu grau de risco, tem os seguintes direitos, a serem exercidos na forma e nas condições descritas neste Capítulo:

I – direito à informação quanto às suas interações com sistemas de IA, de forma acessível, gratuita e de fácil compreensão, inclusive sobre caráter automatizado da interação, exceto nos casos em que se trate de sistemas de IA dedicados única e exclusivamente à cibersegurança e à ciberdefesa, conforme regulamento;

II – direito à privacidade e à proteção de dados pessoais, em especial os direitos dos titulares de dados nos termos da Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais), e da legislação pertinente;

III – direito à não discriminação ilícita ou abusiva e à correção de vieses discriminatórios ilegais ou abusivos, sejam eles diretos ou indiretos.

§ 1º A informação referida no inciso I do caput deste artigo será fornecida com o uso de ícones ou símbolos uniformizados facilmente reconhecíveis, sem prejuízo de outros formatos.

§ 2º Os sistemas de IA que se destinem a grupos vulneráveis deverão, em todas as etapas de seu ciclo de vida, ser transparentes e adotar linguagem simples, clara e apropriada à idade e à capacidade cognitiva, e ser implementados considerando o melhor interesse desses grupos.

§ 3º A utilização de conteúdos de imagem, áudio, voz ou vídeo que retratem ou identifiquem pessoas naturais pelos sistemas de IA deverá respeitar os direitos da personalidade, na forma prevista na Lei nº 10.406, de 10 de janeiro de 2002 (Código Civil), resguardado o direito à liberdade de expressão (NOVO PARÁGRAFO).

A Secretaria de Políticas Digitais optou por seguir com a inclusão do direito à imagem como um novo inciso ao caput do art. 5º. A opção por essa alternativa se deu porque com isso, o direito à imagem é colocado em par com os demais direitos previstos nos incisos já aprovados pelo Senado.

Hipótese de risco excessivo

Na redação do PL 2338/2023 aprovada no Senado, o escopo de sistemas de IA de risco excessivo é de suma importância para definir os sistemas de IA cujo desenvolvimento, implementação e uso são vedados.

A redação atual do artigo do projeto de lei referente a riscos excessivos é limitada no que diz respeito a *deepfakes*, prevendo nesse tema apenas a vedação aos sistemas de IA "com o propósito de" "possibilitar a produção e disseminação ou facilitar a criação de material que caracterize ou represente abuso ou exploração sexual de crianças e adolescentes".

Um primeiro ponto de possível aprimoramento do texto diz respeito à intencionalidade, já que a redação do inciso veda apenas sistemas de IA que tenham os "propósitos" elencados nas alíneas. Um segundo aspecto é o escopo restrito do inciso, que veda apenas material de abuso ou exploração sexual de crianças e adolescentes.

Ainda, o artigo não faz referência a sistemas de IA de produção ou de disseminação de material que caracterize imagens íntimas não consentidas ou outros usos que possam afetar direitos. Considerando o impacto da produção e circulação de imagens de nudez não consentidas por IA, especialmente na vida de mulheres, adolescentes e crianças, o Grupo de Trabalho considerou incluir também essa questão nas vedações. Porém, ante às discussões em torno da legalidade da indústria audiovisual de utilização desses recursos e possíveis benefícios decorrentes para as pessoas profissionais do sexo, optou-se por uma abordagem não proibitiva, mas de identificação obrigatória do usuário para uso dos sistemas para esse fim, que pareceu mais equilibrada. Esse último aspecto passou a ser tratado, então, na seção de IA de Propósito Geral e Generativa.

Dessa forma, as sugestões produzidas pelo GT para o artigo de vedação de desenvolvimento, distribuição e uso de sistemas de IA foram:

Sugestão A - Inserção de um novo inciso, para a vedação clara de IAs que gerem conteúdo sintético de abuso sexual de crianças e adolescentes ainda que esse não seja o propósito do sistema. Também se propõe a inclusão de nova alínea para a vedação de sistemas de IA com propósito discriminatório.

> *Art. 13. São vedados o desenvolvimento, a implementação e o uso de sistemas de IA:*
>
> *I – com o propósito de:*
>
> *e) discriminar ou instigar a discriminação de pessoas ou grupos de pessoas, de forma abusiva ou ilícita, conforme estabelecido na Lei n. 7.716, de 5 de janeiro de 1989.*
>
> *[...]*

V - Que possibilite a criação de conteúdo sintético que:

a) caracterize ou represente criança ou adolescente em atividade sexual ou exibição de nudez de uma criança ou adolescente, ainda que não represente pessoa natural;

Sugestão B - Novo artigo (separação do art. 13): um novo artigo, que seja voltado especificamente a "*deepnudes*" permitiria remover a intencionalidade ("com o propósito de"), sem fazer alterações maiores ao texto aprovado pelo Senado Federal e, assim criar incentivo para que os sistemas de IA não permitam, se forma alguma, a produção de imagens de exploração sexual de crianças e adolescentes.

Art. 13A. É vedado o desenvolvimento e a oferta de sistemas de IA:

I - que produza ou dissemine material que caracterize ou represente abuso ou exploração sexual de crianças e adolescentes, ainda que não represente pessoa natural;

Sugestão C - Alteração de alínea do atual inciso I do art. 13: Essa sugestão manteria a redação do art. 13 do texto aprovado pelo Senado Federal, mas alteraria a alínea "d", para esclarecer que a vedação se aplica ainda que a pessoa representada não seja pessoa natural. Porém, mantém o PROPÓSITO como elemento para a vedação, o que pode diminuir sua eficácia.

Art. 13. São vedados o desenvolvimento, a implementação e o uso de sistemas de IA:

I - com o propósito de:

d) produzir ou disseminar material que caracterize ou represente abuso ou exploração sexual de crianças e adolescentes, ainda que não represente pessoa natural;

A primeira das alternativas listadas acima foi a escolhida pela Secretaria de Políticas Digitais para orientar sua posição, por ela oferecer maior proteção a crianças e adolescentes e contra abuso sexual e práticas discriminatórias.

Proteção de pessoas naturais em IA de geração de conteúdo sintético

O art. 19 do PL 2338/2023 prevê o dever de inclusão de identificador em conteúdos sintéticos gerados por inteligência artificial. No texto aprovado pelo Senado Federal, esse dispositivo está inserido no Capítulo IV - "Da governança dos sistemas de inteligência artificial", na seção II - "Das medidas de governança para sistemas de alto risco", o que leva à conclusão de que apenas sistemas de alto risco teriam o dever de aplicar o identificador. Porém, como os casos de alto risco estão relacionados a hipóteses de aplicação em setores específicos, a IA Generativa de propósito geral, já amplamente disseminada, não seria abarcada. Ou seja, a redação seria ineficaz em boa medida.

Uma primeira sugestão seria, portanto, a alteração da localização do artigo, de forma a passá-lo para a seção V - "Das Medidas de Governança para Sistemas de Inteligência Artificial de Propósito Geral e Generativa". Com isso, o dever de inclusão de identificador valeria para todos os sistemas de IA generativa, não apenas aqueles que fossem de alto risco.

Porém, o Grupo de Trabalho apontou que o dispositivo atual não conta com um parágrafo específico para tratar de conteúdo sintético do tipo *deepfakes*, às quais se aplicariam as regras gerais de identificador. Assim, foi sugerido inserir dispositivos específicos para conteúdos sintéticos que se assemelham à imagem de pessoa natural. O dispositivo atual prevê a inclusão de identificadores associados ao sistema de IA, mas pode ter sua redação aprimorada. Para a inclusão de previsão sobre identificadores sobre o material gerado, uma alternativa seria a inclusão de artigo específico para tal.

> *Art. 19. Quando o sistema de IA for capaz de gerar conteúdo sintético, deverá, considerando o estado da arte do desenvolvimento tecnológico e o contexto de uso, incluir identificador em tais conteúdos para verificação de autenticidade ou de características de sua proveniência, modificações ou transmissão, conforme regulamento*

Sugestão A: Inclusão de novo artigo, referente à necessidade de consentimento para IA generativa que retrate pessoa natural.

> *Art. 33A. Quando o sistema de IA for capaz de gerar conteúdo sintético audiovisual que remeta a imagem, voz ou outros atributos da personalidade de pessoa natural, deverá, considerando o estado da arte do desenvolvimento tecnológico, incluir rótulo ou sinalizador facilmente perceptível por pessoas naturais e identificador legível por máquina contendo informação sobre:*
>
> *I - o consentimento da pessoa, de seus herdeiros ou sucessores para a geração da imagem;*
>
> *II - a identificação do responsável legal pelo conteúdo gerado;*
>
> *§ 1º Quando o sistema possibilitar a produção ou disseminação ou facilitar a criação de material que caracterize imagem de pessoa em atividade sexual ou nua, o identificador deve incluir informação sobre o consentimento, livre, expresso e informado da pessoa retratada;*

§ 2º A criação de conteúdo sintético que utilize a imagem, voz ou qualquer atributo que identifique uma criança ou adolescente, só ocorrerá em acordo com a lei nº 8.069 de 13 de julho de 1990 (Estatuto da Criança e do Adolescente) e dependerá de consentimento livre, expresso e informado de, pelo menos, um dos pais ou do responsável legal, a ser fornecido em conformidade com a Lei nº 13.709, de 2018 (Lei Geral de Proteção de Dados Pessoais)

Sugestão B - Inclusão de parágrafos com ressalva de sátira ou paródia:

Art. XX. Quando o sistema de IA gerar conteúdo sintético, deverá, considerando o estado da arte do desenvolvimento tecnológico ~~e o contexto de uso~~, incluir identificador em tais conteúdos para verificação de autenticidade ou de características de sua proveniência, modificações ou transmissão, conforme regulamento.

§ 1º A presença do identificador previsto no caput não supre outros requisitos de informação e transparência, bem como outros parâmetros a serem definidos em regulamento.

§ 2º A autoridade competente, em colaboração com o Conselho Permanente de Cooperação Regulatória de Inteligência Artificial (Cria), disponibilizará biblioteca de softwares com vistas a facilitar o cumprimento da obrigação de sinalização, idealmente adotando padrão internacional amplamente reconhecido.

§ 3º O uso de conteúdo sintético em obras com finalidade artística, cultural ou de entretenimento poderá, sempre que não representar risco de disseminação de informações falsas, ser sinalizado por meios que não comprometam a utilidade e a qualidade da obra, como nos créditos ou nos metadados associados a tal obra, preservando sua fruição pelo público e seus usos convencionais

§ 4º - Quando o sistema de IA gerar conteúdo sintético que se assemelhe a imagem de pessoa, o desenvolvedor deve incluir no identificador informação sobre o consentimento da pessoa, de seus herdeiros ou sucessores para a geração da imagem; (NOVO PARÁGRAFO)

§ 5º - A obrigação prevista no § 4º não se aplica nos casos em que a pessoa retratada for figura pública e o conteúdo sintético for gerado para fins de sátira ou paródia ou para finalidades críticas ou humorísticas ou para ajustes destinados a melhorar a qualidade da imagem e do som, ressalvada a possibilidade de responsabilização em caso de abuso de direito; (NOVO PARÁGRAFO)

§ 6º Quando o sistema possibilitar a produção ou disseminação ou facilitar a criação de material que caracterize imagem de pessoa em atividade sexual ou nua, o identificador deve incluir informação sobre o consentimento expresso e destacado da pessoa retratada e a identificação do usuário para o sistema de IA; (NOVO PARÁGRAFO)

§ 7º A criação de conteúdo sintético que utilize a imagem, voz ou qualquer atributo que identifique uma criança ou adolescente, dependerá de consentimento prévio, explícito, livre, informado e verificável de, pelo menos, um dos pais ou do responsável legal, a ser fornecido em conformidade com a Lei nº 13.709, de 2018 (Lei Geral de Proteção de Dados Pessoais). (NOVO PARÁGRAFO)

Sugestão C - Inclusão de parágrafo com ressalva de sátira ou paródia

> *Art. XX. Quando o sistema de IA gerar conteúdo sintético, deverá, considerando o estado da arte do desenvolvimento tecnológico ~~e o contexto de uso,~~ incluir identificador em tais conteúdos para verificação de autenticidade ou de características de sua proveniência, modificações ou transmissão, conforme regulamento*
>
> *§ 1º A presença do identificador previsto no caput não supre outros requisitos de informação e transparência, bem como outros parâmetros a serem definidos em regulamento.*
>
> *§ 2º A autoridade competente, em colaboração com o Conselho Permanente de Cooperação Regulatória de Inteligência Artificial (Cria), disponibilizará biblioteca de softwares com vistas a facilitar o cumprimento da obrigação de sinalização, idealmente adotando padrão internacional amplamente reconhecido.*
>
> *§ 3º O uso de conteúdo sintético em obras com finalidade artística, cultural ou de entretenimento poderá, sempre que não representar risco de disseminação de informações falsas, ser sinalizado por meios que não comprometam a utilidade e a qualidade da obra, como nos créditos ou nos metadados associados a tal obra, preservando sua fruição pelo público e seus usos convencionais*
>
> *§ 4º - A geração de conteúdo sintético que se assemelhe a imagem visual ou a voz de pessoa natural requer a obtenção do consentimento da respectiva pessoa, excetuadas hipóteses de exercício protegido da liberdade de expressão, como paródia, sátira ou crítica política. (NOVO PARÁGRAFO)*

A segunda das alternativas ("Sugestão B") listada acima foi a escolhida pela Secretaria de Políticas Digitais para orientar sua posição, por ser baseada na redação já aprovada no Senado Federal para o art. 19 e por ela equilibrar a necessidade de garantir a liberdade de expressão, ao mesmo tempo em que previne abusos e resguarda direitos.

Isso porque estabelece, como regra, a inserção de dados para identificação do conteúdo como sintético e possibilidade de identificação de proveniência. Apenas nos casos mais gravosos, (i) uso de imagem de criança ou adolescente e (ii) geração de imagem sintética de nudez não consentida há requisitos adicionais, quais sejam: para (i) a necessidade de consentimento parental e, no caso de (ii), a identificação do usuário gerador da imagem para o proprietário do sistema, de forma que, ante crime, possa ser identificado. O consentimento que consta da redação é aquele da Lei nº 13.709, de 2018 (Lei Geral de Proteção de Dados Pessoais), que é específico ao tratamento de dados pessoais, mas alternativamente pode se mostrar adequado referenciar o consentimento previsto no art. 20 do Código Civil (lei nº 10.406, de 2002), referente ao direito à imagem.

Proteção da integridade da informação

Ante a preocupação de integrantes do GT quanto ao impacto da geração e disseminação de *deepfakes* para o direito à informação, a integridade da informação e democracia, foram elaboradas alternativas prevendo que as aplicações de Inteligência Artificial devam garantir a inclusão de identificadores nos conteúdos sintéticos, seja com a inclusão de identificador visível para humanos, seja de identificador legível por máquinas. A inclusão de identificador perceptível por pessoas naturais é importante para que a pessoa que visualizar o conteúdo tenha ciência de que se trata de conteúdo sintético. Já a inclusão de identificar legível por máquina é importante para que sistemas automatizados possam realizar operações a partir do dado de que se trata de um conteúdo sintético. Considerando que os conteúdos sintéticos podem ser usados de forma legítima por indústrias para as quais a inclusão de identificar pode ser prejudicial, pode se mostrar oportuno a previsão da possibilidade de dispensa de identificador. Esta flexibilização pode ser importante para usos de IA na indústria publicitária ou artística.

Sugestão A: Inclusão de novo artigo para que todo conteúdo sintético audiovisual deva registrar proveniência e identidade do gerador. Há necessidade de adequação para indústria audiovisual profissional, de forma a que tais obrigações não representem impeditivo a ao uso da IA generativa nessas indústrias.

> *Art. XXA. Quando o sistema de IA gerar conteúdo sintético audiovisual, deverá, considerando o estado da arte do desenvolvimento tecnológico:*
>
> *I - incluir identificador legível por máquina acerca da proveniência, nos termos da regulamentação;*
>
> *II - incluir sinalizador facilmente perceptível por pessoas naturais de que trata-se de conteúdo audiovisual gerado por IA, sempre que o usuário não estiver identificado nos metadados, nos termos de regulamento.*
>
> *Parágrafo único. A autoridade competente determinará os casos de dispensa da obrigação de inclusão de identificador no conteúdo gerado por indústrias específicas.*
>
> *(NOVO ARTIGO)*

A opção da Secretaria foi por incorporar esse dispositivo como art. 31, com aprimoramentos, conforme redação:

> *Art. 31. Sempre que gerarem ou modificarem conteúdo sintético como áudio, imagem ou vídeo as aplicações de IA de propósito geral e generativa deverão incluir identificadores técnicos de proveniência que possibilitem verificar a sua origem e as modificações realizadas no conteúdo, seguindo os padrões definidos em regulamento.*
>
> *§ 1º Para os fins do disposto no caput, os identificadores técnicos deverão ser compostos, cumulativamente, pelos seguintes mecanismos:*
>
> *I - rótulo ou sinalizador facilmente perceptível por pessoas naturais de que se trata de conteúdo gerado por IA, nos termos de regulamento.*

II - identificador legível por máquina acerca da proveniência, desenvolvidas em formatos técnicos interoperáveis, nos termos da regulamentação;

III - assinaturas criptográficas ou outras tecnologias incorporadas ao arquivo que permitam o registro em uma base de proveniência, bem como a posterior verificação de sua origem e integridade pelo público em geral.

§ 2º A presença do identificador previsto no caput não supre outros requisitos de informação e transparência, bem como outros parâmetros a serem definidos em regulamento.

§ 3º A autoridade competente determinará os casos de dispensa da obrigação de inclusão de identificador no conteúdo gerado.

§ 4º Quando o conteúdo sintético gerado incluir a imagem, a voz ou outros atributos da personalidade de uma pessoa natural, a aplicação deverá apresentar informação acerca da possível responsabilidade pelo uso de imagem, voz ou outros atributos de personalidade nos termos do Código Civil, nos termos do regulamento

§ 5º A criação de conteúdo sintético que utilize a imagem, voz ou qualquer atributo que identifique uma criança ou adolescente, só ocorrerá em acordo com lei nº 8.069 de 13 de julho de 1990 (Estatuto da Criança e do Adolescente) e dependerá de consentimento livre, expresso e informado de, pelo menos, um dos pais ou do responsável legal, a ser fornecido em conformidade com a Lei nº 13.709, de 2018 (Lei Geral de Proteção de Dados Pessoais).

Proteção contra *deepnudes*

Considerando o alto nível de dano que imagens e vídeos produzidos por sistemas de IA e que retratam pessoas naturais nuas ou em cenas de sexo, o Grupo de Trabalho considera fundamental aprimoramento da redação do PL 2338/2023 aprovada no Senado Federal. A opção de determinar a vedação do desenvolvimento, distribuição e uso de sistemas de IA para esse fim foi considerada pelo GT. Porém, considerando a possibilidade de usos legítimos, os integrantes desenvolveram proposta que, na avaliação interna, avança em termo de proteção para as cidadãs e cidadãos, sem determinar que as pessoas jurídicas tenham que vedar totalmente tais usos.

Um dos caminhos considerados foi a criação de obrigatoriedade de identificação do usuário criador da imagem ou vídeo ao sistema, de forma que, em caso de ilícito, a autoria seja rapidamente identificada. Isso determina a obrigação de os desenvolvedores de sistemas incluírem essa ferramenta e onera pouco os usuários - uma vez que há possibilidade de usos legítimos.

Sugestão A: Além da mudança de local do art. 19, para que os deveres de sinalização valham também para sistemas de IA generativa que não são de alto risco, e de mudanças à redação dessas obrigações, sugere-se uma regra específica para a geração de *deepfakes* com cenas de nudez ou atos sexuais, as chamadas *"deepnudes".*

Art. AA. Para a geração de conteúdo sintético que retrate pessoa natural em cenas de nudez ou de atos sexuais, os desenvolvedores de aplicações deverão exigir o consentimento, livre, expresso e informado da pessoa retratada, bem como a identificação do responsável legal pelo conteúdo gerado e garantir a possibilidade de revogação do consentimento, nos termos do regulamento.

§1º A autenticação de que trata o caput deverá ser implementada em conformidade com o disposto na Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais).

§2º A declaração acerca do consentimento e a informação sobre a autoria do conteúdo, apurada por meio da autenticação, previstas no caput, deverão constar no registro de proveniência de que trata o art. XX desta lei.

Ante as reflexões desenvolvidas no âmbito do GT, a Secretaria de Políticas Digitais considera importante que, para a geração de conteúdo sintético que retrate pessoa natural em cenas de nudez ou de atos sexuais, os desenvolvedores de aplicações devam exigir ainda, o consentimento, livre, expresso e informado da pessoa retratada, bem como a identificação do responsável legal pelo conteúdo gerado e garantir a possibilidade de revogação do consentimento, nos termos do regulamento, incluindo aprimoramento de redação, pela adição dos dois parágrafos ao artigo.

Alternativamente, se os provedores de aplicação de Internet em que seja possível a criação de *deepnudes* não forem capazes de verificar o consentimento da pessoa retratada para a produção da imagem, tais provedores devem impedir a produção de conteúdo de pessoa identificada ou identificavel em cenas de nudez ou sexo. No caso de verificação de identidade do usuário e o pedido de geração de *deepnude* seja de sua própria imagem, o provedor poderia permitir a geração da imagem.

Avaliação Preliminar para IA generativa

O art. 29 da redação dada ao projeto de lei na versão aprovada no Senado Federal prevê que os desenvolvedores de sistema de IA generativa, o que abarca sistemas de IA capazes de produzir *deepfakes,* realizem avaliação preliminar para identificação dos riscos esperados pelo uso. No entanto, cabe avaliar a possibilidade de aprimorar a redação do artigo para que haja obrigações específicas para sistemas de IA capazes de gerar *deepfakes.*

Cabe avaliar ainda a inclusão de artigo que também preveja a mitigação de risco de sistema de IA generativa anteriormente à sua disponibilização no mercado brasileiro. A inserção desse dispositivo ao projeto de lei seria importante para o enfrentamento aos possíveis danos de *deepfakes* a direitos fundamentais e à integridade da informação.

Sugestão A – Desmembramento de parágrafo único em 5 incisos: essa alternativa amplia o escopo do dever de avaliação preliminar de risco pelos fornecedores de sistemas de IA de propósito geral e generativa, de forma a incluir o dever de identificar também riscos associados à possibilidade de o sistema produzir conteúdo sintético.

Art. 29. O desenvolvedor de sistemas de IA de propósito geral e generativa deverá realizar, além da documentação pertinente sobre o desenvolvimento do sistema, sua avaliação preliminar de risco, a fim de identificar os níveis de risco esperados, inclusive potencial risco sistêmico e incluindo, pelo menos:

~~*Parágrafo único. A avaliação preliminar deverá considerar as finalidades de uso razoavelmente esperadas e os critérios previstos, nos termos da Seção III do Capítulo III desta Lei.*~~

I - a lista das finalidades de uso razoavelmente esperadas e a possibilidade de enquadramento nos critérios previstos para consideração enquanto sistema de alto risco;

II - a possibilidade de o sistema produzir conteúdo sintético assemelhado à realidade;

III - os riscos de reprodução de vieses discriminatórios,

IV - a possibilidade de o sistema produzir conteúdo sintético a partir de referências de pessoa natural;

V - as medidas adotadas para mitigar os riscos associados aos usos previstos nos incisos II , III e IV, indicando sua proporcionalidade e razoabilidade.

Sugestão B - Inserir artigo 31 na Seção V, prevendo dever de identificação e mitigação de riscos de IA generativa.

Art. 31. O desenvolvedor de um sistema de IA generativa deve, antes de disponibilizar no mercado para fins comerciais, realizar e documentar ações para identificar riscos razoavelmente previsíveis e mitigá-los, especialmente quanto a:

I - proteção de direitos fundamentais, especialmente de crianças e adolescentes;

II - proteção da integridade da informação,

III - garantia da liberdade de expressão e do acesso à informação.

Parágrafo único. O desenvolvedor deverá tornar disponível, sempre que solicitado pelos agentes do SIA, no âmbito de processo administrativo específico, material comprobatório das medidas mencionadas no caput. ***(NOVO ARTIGO)***

Ante os debates realizados no âmbito do GT, e as propostas desenvolvidas para lidar com o fenômeno das *deepfakes* em outros pontos do texto, destacadamente os aprimoramentos para transparência quanto a conteúdo gerado por IA, inclusão de dados de proeminência nos produtos audiovisuais gerados e obrigação de identificação dos usuários em caso de *deepnudes*, a Secretaria de Políticas Digitais optou por sugerir alterações ao art. 29 do texto aprovado no Senado, de forma a preservar a avaliação preliminar e incluir os incisos mencionados na "sugestão A" quanto à lista das finalidades de uso razoavelmente esperadas, à reprodução de vieses discriminatórios e à mitigação de riscos. No que diz respeito aos incisos II e IV, optou-se por restringir seu escopo e aglutina-los, em um inciso que preveja o dever a avaliação preliminar abarcar "o risco de geração de conteúdo fraudulento que utilize, sem autorização, a identidade de pessoas públicas, contas ou marcas, que possa gerar vantagem ilícita, em prejuízo alheio, com indução de terceiros a erro".

Artigo específico sobre *deepfakes*

Alternativamente à inclusão de diversos dispositivos ao longo do PL que permitam o regramento de *deepfakes*, também seria possível incluir artigo voltado especificamente para isso.

Sugestão A - Inclusão de dever específico a *deepfakes* sobre a identificação do conteúdo gerado.

> *Art. XX. Os responsáveis pela implantação de um sistema de IA que gere ou manipule conteúdos de imagem, áudio ou vídeo que constituam um conteúdo sintético assemelhado à realidade devem revelar, de forma compreensível e clara, que os conteúdos foram artificialmente gerados ou manipulados, ainda quando se refiram a fins particulares e não econômicos.*
>
> *§1º A obrigação prevista no caput não se aplica se a utilização for autorizada por lei para detectar, prevenir, investigar ou reprimir infrações penais.*
>
> *§2º Sempre que os conteúdos referidos no caput façam parte de um programa ou obra de natureza manifestamente artística, criativa, satírica, ficcional ou análoga, as obrigações de transparência estabelecidas no presente artigo limitam-se à divulgação da existência desses conteúdos gerados ou manipulados, de uma forma adequada que não prejudique a exibição ou a fruição da obra.*
>
> *§3º Os sujeitos referidos no caput deverão adotar medidas técnicas capazes de inserir no conteúdo artificialmente criado marcadores que possam ser usados por outros sistemas de inteligência artificial para reconhecer sua natureza artificial.*
>
> *§4º Os provedores de aplicação, especialmente plataformas de redes sociais e mensageria privada instantânea, deverão adotar medidas técnicas para detectar e identificar de forma explícita aos usuários conteúdos criados por inteligência artificial.*
>
> *§5º A recriação digital e a exploração da imagem de pessoas falecidas deverão obedecer às seguintes condições:*
>
> *I – ressalvadas as exceções previstas pela legislação, será exigida a obtenção prévia e expressa de autorização, para fins específicos e determinados, da pessoa em vida ou, na sua falta, dos legitimados previstos pelo parágrafo único do artigo 12 da Lei nº. 10.406/2002;*
>
> *II – respeito à imagem-atributo construída em vida pela pessoa falecida.*
>
> *III - terá o Ministério Público legitimidade para tutelar violações à imagem-atributo das pessoas falecidas em situações em que haja evidente violação por parte dos herdeiros, considerado o interesse público na manutenção da integridade do legado dessas pessoas.*

Sugestão B - Inclusão de dever específico a *deepfakes* sobre o respeito ao direito à imagem.

> *Art. X. A utilização de conteúdos de imagem, áudio ou vídeo que retratem ou identifiquem pessoas naturais, direta ou indiretamente, para o treinamento ou funcionamento de sistemas de inteligência artificial deverá respeitar os direitos da personalidade, em especial os direitos à imagem, à voz e à identidade pessoal, na forma prevista pela legislação.*
>
> *§1º Ressalvadas as exceções previstas pela legislação, é vedada a utilização dos conteúdos referidos no caput, ainda que para fins de treinamento do sistema de inteligência artificial, sem autorização da pessoa natural a quem se refiram, observadas as regras de direito autoral e de proteção de dados pertinentes.*
>
> *§2º Nos termos do parágrafo anterior, é vedada a coleta aleatória e indiscriminada dos conteúdos referidos no caput a partir da Internet, especialmente quando tal coleta seja realizada por meio de mecanismos automatizados e tenha por finalidade criar ou expandir bases de dados de reconhecimento facial ou inferência de emoções, observadas as regras de proteção de dados pertinentes.*
>
> *§3º A utilização dos conteúdos referidos no caput deverá se dar para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de utilização posterior de forma incompatível com essas finalidades, aplicando-se, no que couber, as regras previstas no artigo 6º da Lei nº. 13.709/2018.*
>
> *§4º O legislador deverá disciplinar as hipóteses de licenciamento comercial dos conteúdos referidos pelo caput para exploração em obras que envolvam interpretações ou execuções, sendo vedadas, desde logo, as cláusulas contratuais que prevejam a possibilidade de recriação artificial por tempo indefinido e para finalidades indeterminadas.*
>
> *§5º As presentes disposições aplicam-se, no que couber, às pessoas jurídicas, na forma do disposto no artigo 52 da Lei nº. 10.406/2002.*

A Secretaria de Políticas Digitais entendeu que a abordagem de fazer alterações ao longo do texto, que dialogassem com a versão já aprovada no Senado, seria melhor do que concentrar as regras de *deepfakes* em um único artigo. Assim, nenhuma das alternativas acima foi levada à frente como posição interna para o debate interministerial, mas elas ainda podem ser alternativas para a disciplina de *deepfakes*, seja no PL 2338/2023 ou em legislação específica.

2 Conclusões

As *deepfakes* colocam em evidência um conjunto de desafios múltiplos, e o enfrentamento a eles é fundamental para assegurar que a IA possa avançar sem colocar em risco o exercício de direitos. Para que as inovações em torno da IA, e em particular da IA generativa, avancem sem prejuízo a direitos, é necessário que a regulação da IA conte com mecanismos jurídicos específicos para a proteção de direitos em face a *deepfakes*.

Embora o texto do PL 2338/2023 aprovado pelo Senado apresente dispositivos essenciais para a regulação à IA, e conte com uma seção específica para IA generativa, o projeto ainda não conta – inclusive por conta da velocidade de avanço da tecnologia - com dispositivos específicos suficientes para proteger os direitos em face às *deepfakes*, conforme avaliação dos e das integrantes do GT.

Nesse sentido, o PL 2338/2023 desponta como uma oportunidade concreta para incorporar dispositivos que fortaleçam a capacidade regulatória do país frente ao fenômeno das *deepfakes*. Ajustes pontuais podem contribuir para suprir lacunas emergentes e ampliar a proteção jurídica.

Apesar de o trabalho do GT ter ficado concentrado na análise do PL 2338/2023, lacunas e possíveis aprimoramentos, resta evidente a urgência de soluções que transcendam a esfera legal — seja por meio de políticas públicas, cooperação internacional ou desenvolvimento de ferramentas técnicas. Somente uma combinação integrada de estratégias poderá enfrentar de modo efetivo os impactos sociais e democráticos envolvidos no fenômeno das *deepfakes*.

A Secretaria de Políticas Digitais, por fim, agradece a todos que integraram o Grupo de Trabalho e registra seus debates e resultados como parte da contribuição da Secretaria de Comunicação Social da Presidência da República a esse relevante tema.

Secretaria de Políticas Digitais
Secretaria de Comunicação Social
Presidência da República